数理解析研究所講究録
312巻 1977年 115-124

時系列推定における高次の漸近有効性について

竹内　啓

§1

時系列における母数推定の理論は、一面から見れば過去の数十年間に大いに進歩したが、しかしなお最適推定論 optimum estimation の観点からは、極めて不十分なところがあるといわざるを得ない。

いま時系列データ $X_1(\omega)$, $X_2(\omega)$, $\cdots$ $X_t(\omega)$ $\cdots$ が与えられたものとしよう。このとき推定論の問題を明確に定式化するためには、次のものを明確に定義しなければならない。

1. $\Omega=\{\omega\}$ 上で定義された確率測度の族 $\{P_\omega^\theta\}$ $\theta\in\Theta$
2. 推定すべき母数（実、または実ベクトル）$\gamma=g(\theta)$
3. 考慮される推定量のクラスと、その「よさ」の基準

時系列における推定論の固有の問題は、上記の1～3が明確に定義されたことを前提にした上で、最適な推定量を求めること、或いは特定の推定量の相対的な効率を求めることである。従って時に問題にされる推定量の計算手続きの問題は推定論の固有の問題ではない。勿論いわゆる「簡便法」の相対効率は問題となり得るけれども。

時系列解析の書物において、上記の1～3の点、とくに1と3に関して明確な規定が欠けていることが、問題を不明確

にしていることが少くない。例えばいわゆる「スペクトル密度の推定」において、考慮の対象になっている分布の集合がどのようなものであるのか、より具体的にいえば「真の」密度関数がどのような集合に属するものであるかが明らかでないことが、いろいろな推定法の比較を直観的な議論の水準にとどめている。

実際時系列解析において、いわゆる「Non-parametric」すなわち具体的な分布形を仮定しないモデルについて、厳密な推定論を展開することは、抽象的な測度論のレベルでも、また推定法の比較という観点からも、著しく困難であるように思われる。推定論の観点からは「parametric」な、より具体的にいえばガウス過程を中心としたモデルだけがmanageableであるといわねばならない。しかし現実のデータがこのような特定の分布形を持つ分布に厳密に従うということは保証されないかもしれない。このような場合を処理するためにはモデルからの乖離によって影響されるところの少いような手法を求めるrobustnessの観点を導入しなければならない。

しかしここではrobustnessの問題はとりあげないことにしたい。すなわち「正しい」parametric model が与えられているものとしよう。更に分布を規定する母数θ自体が

有限次元の実ベクトルで表されるものと仮定する。このときには推定すべき母数を θ 自体と考えてよい。

ところで時系列モデルにおいては、厳密な小標本理論を展開することは不可能に近い。例えば最小分散不偏推定量が存在するような場合はほとんどない。そもそも不偏推定量の存在自体が疑問である場合が多い。もっともその不存在の証明は一つの興味ある問題であるが。

従って理論的に期待し得る結果は、ほとんど漸近的なものに限られるであろう。実際これまでに得られている結果も、ほとんど漸近理論の範囲内のもののみである。しかしこの分野にもなおいくつかの問題がある。

1 推定量の漸近的最適性 asymptotic optimalityの定義と厳密な定式化。

2. 漸近分布への収束の速さと高次の漸近最適性 asymptotic optimality of higher orderの問題

1に関しても、しばしば最尤推定量の漸近最適性などが厳密な裏付けなしにいわれていることが多い。或いは独立同一分布の観測値からなる標本についての結果が、そのまま援用されていることも少くない。いうまでもなくこのことは理論的に妥当でないが、しかしながら時系列解析のふつうのモデルに関する限り、最尤推定量の漸近正規性、漸近有効性等

について，結果的には誤まった結論に導びかれることはないようである．標準的な漸近理論の方法（竹内啓「統計的推定の漸近理論」参照）が，中心極限定理の援用などにおいてしかるべき注意さえ払えば，ここでも適用できる。

しかしながら時系列モデルにおいては，極限分布への収束はあまり速くないのが通例である．また同じ理由によることであるが，漸近的には同等になる二つの推定量が，現実に有意に異る値を示すこともある。従って高次の漸近分布についての立ち入った議論が必要である．

しかしこの点に関しては，独立同一分布の場合についてもまだ理論は完成していない。またこれまでに得られた理論を時系列モデルに適用するについては，独立でない確率変数の和の分布のEdgeworth展開を理論的に基礎づけなければならないが，それもあまり容易ではない。

以下においては，最近になって明らかとなった高次の漸近有効性に関する「拡張された正則推定量」の性質についてのべ，その時系列モデルへの応用について論じたい。それによって問題を考察するための手がかりを得ることが，この文の主要な目的である．

§ 2

最初に筆者によって最近得られた結果を紹介する．

$X_1, X_2 \cdots X_n$ …が互いに独立に次のような密度を持つ指数型分布に従うとする。

$$f(x, \theta) = c(\theta) h(x) \exp \sum_{i=1}^{p} s_i(\theta) t_i(x)$$

ここに t_i は実数値をとる関数で、それらは互いに線形独立とする。また $\theta \in \Theta \subset R^q$ は q 次元実母数で $q < p$、$\theta \in \Theta$ に対し $(s_1(\theta), \cdots s_p(\theta))$ は $(s_1 \cdots s_p)$ のとり得る値の自然な範囲の内点になっているとする。また s_i は θ に関し適当な次数まで連続微分可能であると仮定する。

上記の分布に対して十分統計量は

$$T = (T_1 \cdots T_p) \qquad T_i = \frac{1}{n}\sum_{j=1}^{n} t_i(X_j)$$

で与えられる。T の分布は $n \geqq p$ のとき Lebesgue 測度に関し絶対連続であると仮定する。そうすると T はすべての次数のモーメントを持ち、かつその分布の密度関数を Edgeworth 展開することができる。

θ の「拡張された正則推定量」extended regular estimator とは次のような形に表される推定量をいう。

$$\hat{\theta} = g(T_1 \cdots T_p) + \frac{1}{n} h_1(T_1 \cdots T_p) + \cdots + \frac{1}{n^k} h_k(T_1 \cdots T_p)$$

ここに $g, h_1 \cdots h_k$ はすべて n をふくまない関数で g は $2k-1$ 回、h_j は $k-j$ 回それぞれ連続微分可能と仮定する。このとき、$\hat{\theta}$ の Taylor 展開が可能であって、それにもとづいてその漸近分布の密度関数の Edgeworth 展開を与えることも

ができる。

ここで $k=1$ とすると、$\hat{\theta}$ の密度の $1/n$ の order までの漸近展開が得られる。そうして $\sqrt{n}(\hat{\theta}-\theta)$ の漸近 cumulant が

$$E(\sqrt{n}(\hat{\theta}_\alpha-\theta_\alpha)) = \mu_\alpha/\sqrt{n} + o(1/n)$$

$$E(\sqrt{n}(\hat{\theta}_\alpha-\theta_\alpha)(\hat{\theta}_\beta-\theta_\beta)) = \sigma_{\alpha\beta} + Q_{\alpha\beta}/n + o(1/n)$$

$$E(\sqrt{n}(\hat{\theta}_\alpha-\theta_\alpha)(\hat{\theta}_\beta-\theta_\beta)(\hat{\theta}_\gamma-\theta_\gamma)) = \beta_{\alpha\beta\gamma}/\sqrt{n} + o(1/n)$$

$$E(\sqrt{n}(\hat{\theta}_\alpha-\theta_\alpha)(\hat{\theta}_\beta-\theta_\beta)(\hat{\theta}_\gamma-\theta_\gamma)(\hat{\theta}_\delta-\theta_\delta)) - \sigma_{\alpha\beta}\sigma_{\gamma\delta} - \sigma_{\alpha\gamma}\sigma_{\beta\delta}$$
$$- \sigma_{\alpha\delta}\sigma_{\beta\gamma} = \gamma_{\alpha\beta\gamma\delta}/n + o(1/n),$$

という形で与えられる。

ここで $\Sigma = (\sigma_{\alpha\beta}) \geq I^{-1}$ (I は情報行列) という関係がなり立ち、かつ等号は $\hat{\theta}$ が最尤推定量 或いはそれと漸近的に同等なBAN推定量の場合に限ってなり立つ。

更にすべてのBAN推定量について、$\beta_{\alpha\beta\gamma}$ および $\gamma_{\alpha\beta\gamma\delta}$ は等しいことが示される 従って漸近分布の $1/n$ の order までの展開において、拡張されたBAN推定量のクラスの中で一定でないのは μ_α および $Q_{\alpha\beta}$ のみである。ところで μ_α は $\hat{\theta}$ について高次の漸近不偏性、高次の漸近中央値不偏性、或いは高次のモード不偏性のいずれかを仮定するならば、$\beta_{\alpha\beta\gamma}$ から定められてしまう。従ってこのような何らかの意味での高次の正則性をみたすクラスの中では $Q_{\alpha\beta}$ のみが問題となる。ここで次のことがなり立つことが証明される。

$$\sum_\alpha\sum_\beta Q_{\alpha\beta} t_\alpha t_\beta \geq \sum_\alpha\sum_\beta Q^*_{\alpha\beta} t_\alpha t_\beta \qquad \forall t_\alpha$$

ただし $Q^*_{\alpha\beta}$ は（拡張された）最尤推定量に対応する値である。このことから次のことがいえる。$\hat{\theta}^*$を最尤推定量を高次の漸近正則性をみたすように補正した推定量　$\hat{\theta}$と同じ漸近正則性をみたす（拡張された）正則推定量とすると、原点をふくむ任意の凸集合 C に対して

$$\varliminf_{n\to\infty} n\left[\Pr\{\sqrt{n}(\hat{\theta}^*-\theta)\in C\}-\Pr\{\sqrt{n}(\hat{\theta}-\theta)\in C\}\right]\geq 0$$

がなり立つ。或いは $L(u)$ を

$L(0)=0 \quad L(u)>0$　すべての $a>0$ に対して

集合 $\{u \mid L(u)\leq a\}$ は凸，$L(u)$ は有界

をみたす関数とすると

$$\varlimsup_{n\to\infty} n\left[E\{L(\sqrt{n}(\hat{\theta}^*-\theta))\}-E\{L(\sqrt{n}(\hat{\theta}-\theta))\}\right]\leq 0$$

がなり立つ。

すなわち（拡張された）最尤推定量は n^{-1}の order までの、いいかえれば3次の漸近有効性を持つことがいえる。

§3

以上の結果は、実は直接に拡張可能であって、時系列モデルにも一部応用できるものである。

すなわち前節の結果を得るための本質的な前提は、有限次元の十分統計量が存在し、かつその密度関数の Edgeworth 展開が可能であるということであって、拡張された正則推定

量の漸近モーメントに関する結果は、これから直接に導びかれるのである。

従って例えば $X_1, X_2, \cdots X_n \cdots$ がすべて互いに独立でその密度関数が

$$f(x_j, \theta) = c_j(\theta) h_j(x_j) \exp \sum_i s_i(\theta) t_{ij}(x_j)$$

という形に表されるとき、もしある関数 $\bar{c}(\theta)$ が存在して

$$\frac{1}{n}\left(\frac{d}{d\theta}\right)^m \sum \log c_j(\theta) \to \frac{d^m}{d\theta^m}\bar{c}(\theta)$$

$$m = 1, 2 \cdots$$

がなり立つならば、前節の結果が成立することは直ちに示される。すなわちこの場合でも（拡張された）最尤推定量は3次の漸近有効性を持っている。またその漸近モーメントは尤度関数の導関数（高次のものもふくめて）の同時モーメントから定められる。

時系列解析の場合に、有限次元の十分統計量が存在するならば、ほとんど同じ結果がなり立つ。例えば自己回帰モデル

$$x_{t+k} = \beta_1 x_{t+k-1} + \beta_2 x_{t+k-2} + \cdots + \beta_k x_t + \varepsilon_t$$

$$\varepsilon_t \sim N(0, \sigma^2)$$

において $x_1 \cdots x_T$ の同時密度関数は

$$\left(\frac{1}{\sqrt{2\pi}\,\sigma}\right)^{T-k} \exp\left\{-\frac{1}{2\sigma^2}\sum_{t=k+1}^{T}(x_t - \beta_1 x_{t-1} - \cdots - \beta_k x_{t-k})^2\right\}$$

$$\times f_0(x_1 \cdots x_k)$$

である。ただし $f_0(x_1 \cdots x_k)$ は"初期値" $x_1 \cdots x_k$ の同時密度密度関数である。そうすると十分統計量は

$$T_0 = \sum_{t=k+1}^{T} x_t^2/(T-k),\ T_1 = \sum_{t=k+1}^{T} x_t x_{t-1}/(T-k) \cdots T_k = \sum_{t=k+1}^{T} x_t x_{t-k}/(T-k)$$

および $x_1 \cdots x_k;\ x_T,\ x_{T-1},\ \cdots x_{T-k+1}$

で与えられることがわかる。(このことは"初期値"の分布についての仮定とは無関係になり立つ)

ここで $T_0, T_1 \cdots$ の同時分布については、Edgeworth 展開が可能であるから、これまでの場合と同様に扱うことができるが、$x_1 \cdots x_k,\ x_T, \cdots x_{T-k+1}$ についてはその持つ情報量は一定であって、T とともに増加することはない。従ってそれを推定量の中にとり入れるについては、特殊の扱いが必要になる。一つの方法は推定量のクラスとして

$$\hat{\beta} = g(T_0\ T_1 \cdots T_k) + \frac{1}{T} h(x_1 \cdots x_k,\ x_T \cdots x_{T-k+1}; T_0 \cdots T_k)$$

という形のものを考えることである。この場合も問題は残されているが本質的な困難はない。

有限次元の十分統計量が存在しない場合、例えば自己回帰・移動平均(ARMA)モデルの場合などには困難は大きい。一つの方法は、十分次元の多くの統計量

$$T_0 = \sum x_t^2/T,\quad T_1 = \sum x_t x_{t-1}/(T-1) \cdots \quad T_r = \sum x_t x_{t-r}/(T-r)$$

をとって、それらにもとづく(拡張された)正則推定量を考え、まず r を固定してその漸近分布を考え、次に r を無限に

大きくすることであろう。Iを固定すれば"補正された"最大推定量が2次の漸近有効性を持つことは　これまでと同様の議論によって示すことができるから、次にIとしてTに応じて適当な速さで増加させればよいであろう。ただしこのような点についての吟味はまだ今後の課題である。